UNITED STATES DISTRICT COURT
DISTRICT OF NEW JERSEY

IQVIA INC. and IMS SOFTWARE SERVICES,)
LTD.,)
)
)
      Plaintiffs –)
      Counterclaim Defendants,)
)
   v.)
)
VEEVA SYSTEMS INC.,)
)
      Defendant –)
      Counterclaim Plaintiff.)
)
)
)
)

Case No.: 2:17-cv-00177-CCC-MF

Hon. Mark Falk

Document Electronically Filed

STIPULATED ORDER
REGARDING TECHNOLOGY ASSISTED REVIEW

The parties agree and respectfully request that the Court order that the procedures outlined herein shall govern the use of technology assisted review ("TAR") for electronically stored information ("ESI") in this litigation.

A.   **Definitions:**

1.   "Coding" means the process of tagging Relevant and Non-Relevant documents.

2.   "Confidence Interval" means the range estimated to contain the true value of a statistical measure (such as recall or precision) within a specific confidence level.

3.   "Confidence Level" means the probability that the confidence interval contains the true value of a statistical measure (such as recall or precision).

4.   "Control Set" means a sample of documents coded at the outset of TAR that may be used to measure the effectiveness of the machine-learning algorithm at various stages of training, and to determine when training may cease.

5.   "Elusion" means the fraction of documents designated as Non-Relevant by TAR, that are in fact Relevant.

6.   "Precision rate" means the percentage of actually Relevant documents out of all documents the TAR algorithm designates Relevant.

7.   "Recall rate" means the percentage of actually Relevant documents in a collection that, through training, the TAR algorithm designates as Relevant.

8.   "Relevant Documents" means documents that are relevant to the claims and defenses at issue in this litigation.

9.   "Training Set" means the set of documents used to train the TAR machine-learning algorithm.

B.   Technology Assisted Review. The Parties will use TAR to aid review of documents before production. Under TAR, a machine-learning algorithm assesses the likely relevance of a corpus of documents based on manual relevance determinations applied to a subset of that corpus.

C.   Data Sources.

1.   The Parties will use TAR for all documents collected from agreed-upon Custodians, aside from the following exceptions:

(a)   Text messages, spreadsheets larger than 2mb, and documents collected from table-based non-custodial databases will not be subject to TAR — the Parties will exchange search terms for use with these document types. In addition, the following file types will be excluded from TAR: ZIP/JAR files, image files, technical drawings, database files, and audio and video files. The Parties will meet and confer if a Party wishes to exclude additional types of documents from TAR.

(b)   Documents containing substantive non-English language content may be excluded from TAR, depending on the number and importance of documents subject to review in that language. The Parties will meet and confer as to the languages that will be subject to review and whether manual review, search term review, or predictive coding will be used for that language. In all cases, the Parties will use an attorney fluent in a language to execute the review plan for documents in that language.

(c)   Prior to the Parties' TAR review process, the Parties may propose to exclude documents from their document collection that were sent to or received from email domains that indicate such documents are clearly Non-Relevant (e.g., personal travel or shopping websites). The Parties reserve the right to object to the opposing Party's exclusion of any particular domain on the basis that documents sent to or received from that domain may be Relevant.

3

(d)   Prior to the Parties' TAR review process, the Parties may propose to exclude computer/system-generated documents from their collection that are Non-Relevant to the claims or defenses in this case. Any proposal to cull such documents will be accompanied by a sample of such documents along with complete or sample metadata. The Parties reserve the right to object to the opposing Party's exclusion of any particular such document (or type of document) on the basis that such documents are Relevant.

The use of the term "documents" in the rest of this Protocol shall mean documents other than those exceptions described in C(1)(a)-(d).

2.   The Parties have agreed to an initial group of 25 Custodians whose documents will be collected and subject to TAR. The Parties reserve the right to identify additional Custodians for TAR at a later date, and likewise reserve the right to object on grounds of relevance or burden to the other side's identification of additional Custodians. The identity and number of additional Custodians, if any, shall be determined by further conferral between the Parties and, only if necessary, resolution by the Court.

3.   All of a Custodian's documents over which a Party has possession, custody, or control will be included in TAR. This includes without limitation documents from any entity acquired by a Party, so long as that Party has possession, custody, or control of a Custodian's documents from a Custodian's employment with that entity.

4.   The default date range for inclusion of Custodians' documents in TAR is January 1, 2012 to January 10, 2017. The Parties reserve the right to request that a

4

particular Custodian's documents from January 1, 2010 to December 31, 2011, or

January 11, 2017 to the present (on an ongoing basis) also be subject to TAR. Any such

request will include a short explanation of the reason the Custodian's documents from

outside the default date range are Relevant to the claims and defenses in this case. The

Parties reserve the right to object to any such request, and any objection will be resolved

by meeting and conferral or, only if necessary, the Court.

D.   TAR Training and Standards.

1.   The Parties' initial TAR training sets will be randomly selected from the Parties'

collection of documents from the initial 25 Custodians identified for collection.

2.   The Parties will train their TAR algorithm by coding training set documents as

Relevant or Non-Relevant to the claims and defenses in this case. Coding shall be

conducted by attorneys at the level of Associate and above who are familiar with the

Antitrust, Trade Secret, Lanham Act, and state law claims and defenses of both Parties.

3.   The Parties will code training sets until they reach a 90% recall rate with a 2.5%

confidence interval at a 95% confidence level. After reaching those standards, the Parties

will meet and confer as to the appropriate precision rate to be reached by each Party. If a

Party comes to believe that these recall rate or confidence measures create undue burden,

the Party may request to meet and confer to discuss these standards.

E.   Production of Documents. The Parties will produce documents designated as Relevant

by TAR unless an attorney reviewer identifies them as privileged or Non-Relevant. Documents

identified as privileged will be treated in accordance with the terms of this Court's Discovery Confidentiality Order and Electronically Stored Information Order.

F.   Search Terms and Verification.

1.   After initial productions of documents and at a date established in accordance with further meetings and conferrals, the Parties will exchange search terms to be run on all documents designated as Non-Relevant by the TAR algorithm or attorney review of the opposing Party. The default number of search terms will be ten or less, but the Parties may meet and confer to vary that number depending on the specificity and reasonableness of search terms submitted.

2.   The search terms will be tailored to particular issues. Indiscriminate terms will be inappropriate unless combined with narrowing search criteria that sufficiently reduce burden. A conjunctive combination of multiple words or phrases (e.g., "computer" and "system") narrows the search and shall count as a single search term. A disjunctive combination of multiple words or phrases (e.g., "computer" or "system") broadens the search, and thus each word or phrase shall count as a separate search term unless 1) the disjunctive combination is limited by other parameters that tailor the term to a specific issue, or 2) the terms in the disjunctive combination are variants of the same word. Use of narrowing search criteria (e.g., "and," "but not," "w/x") is encouraged.

3.   The Parties will, upon the request of the opposing Party, run an initial list of search terms against all documents previously designated Non-Relevant and provide the

6

opposing Party with the number of documents (both inclusive and exclusive of document family members) hit by each search term.

4.   Documents responsive to the Parties' final list of search terms will be reviewed (or re-reviewed) by an attorney and produced if Relevant and non-privileged. Separately, each Party will use the documents responsive to the opposing Party's search terms to further train its TAR machine-learning algorithm. If that additional training causes the TAR algorithm to mark additional documents as Relevant, then those documents will also be produced unless determined to be privileged or Non-Relevant by a human reviewer.

G.   **Measurements and Documentation.** The following information will be captured, documented, and shared by the Parties for each set of Custodians:

1.   The total number of documents, both inclusive and exclusive of document family members, collected from the set of Custodians.

2.   A brief description of any action taken to cull the collected population prior to TAR (e.g., "Documents received from the following domains were identified and culled: espn.com, expedia.com . . . .").

3.   The total number of documents subject to TAR, both inclusive and exclusive of document family members.

4.   The total number and percentage of documents designated Relevant by TAR, both inclusive and exclusive of their Non-Relevant family members.

5.   The total number and percentage of documents subject to attorney review, both inclusive and exclusive of family members. In addition, a brief explanation of any action taken, other than attorney review, to cull Non-Relevant documents (e.g., "Documents from the following domains were deemed automatically Relevant/Non-Relevant based on the following metadata fields: . . . .").

6.   The total number and percentage of documents deemed Non-Relevant during attorney review, both inclusive and exclusive of family members.

H.   Reservations of Rights. The Parties reserve the right to move to amend this TAR Protocol, or seek modification through meetings and conferral. The Parties agree that the stipulations set forth in this Protocol shall not limit their ability to request discrete documents or document categories that may not belong to a particular agreed-upon Custodian or data source. Nothing in this Protocol shall be construed to waive the Parties' rights to request documents or to object to particular discovery requests under the Federal Rules of Civil Procedure. The Parties will confer in good faith upon any such requests or objections. If the Parties cannot reach agreement, any dispute shall be presented to the Court in accordance with the Federal Rules of Civil Procedure, Local Rules, and orders of the Court.

SO ORDERED:

Dated: January 5, 2018

Hon. Mark E. Falk
United States Magistrate Judge

8